

REMARKS

Claims 1, 5-13, 26, 28, 29 and 33 remain in the application. Claims 2-4, 14-25, 27, 30-32 and 34-36 have been cancelled, claim 33 is original, claims 6, 7 and 28 have been previously presented, and claims 1, 5, 8-13, 26 and 29 are currently amended in order to more clearly define the invention. The Examiner has rejected all of the pending claims (1, 5-13, 26, 28, 29 and 33) under 35 U.S.C. § 103(a) as being unpatentable over Bartosik et al. (US 672519) in view of Papineni et al. (U.S. Patent No. 6246981). This rejection is respectfully traversed, and reconsideration is requested in view of the foregoing amendments, and following remarks.

Claim 1 now recites a speech recognition system comprising: a querying device for posing at least one query to a respondent over a telephone; a speech recognition device which receives an audio response from the respondent over the telephone and conducts a speaker-independent speech recognition analysis of said audio response to automatically produce a corresponding text response; a storage device for recording and storing said audio response as it is received by said speech recognition device; an accuracy determination device for automatically comparing said text response to a text set of expected responses and determining whether said text response corresponds to one of said expected responses; wherein if said accuracy determination device determines that said text response does not correspond to one of said expected responses within a predetermined accuracy confidence parameter, said accuracy determination device flags said audio response so as to produce a flagged audio response for further review by a human operator; and a human interface device for enabling the human operator to hear the flagged audio response and review the corresponding text response for the flagged audio response to determine the actual text response for the flagged audio response, either by selecting from a pre-determined list of text responses or typing the actual text response if no such match exists in the pre-determined list of text responses. Claim 1 has been amended to recite that a querying device poses at least one query to a respondent over a telephone; and that the speech recognition device which receives an audio response from the respondent over the telephone, conducts a speaker-independent speech recognition analysis of the audio response to automatically produce a corresponding text response.

BST99 1482066-1.057622.0045

Further, the storage device records and stores the audio response as it is received by the speech recognition device. An accuracy determination device automatically compares the text response to a text set of expected responses and determines whether the text response corresponds to one of said expected responses. Similar changes have been made to other claims.

Thus, the present application claims a system/method for selective human correction or checking of calls that are handled by a speaker-independent speech recognizer. To do this, the caller/respondent's utterances are simultaneously processed with a (automatic, computer) speech recognizer, AND recorded so a human can check it after the call, if necessary. If the confidence parameter for a particular utterance is above a certain threshold, it is not necessary for a human to check the results for accuracy. It is assumed the speech recognizer got it right (and can delete the recording). If the confidence parameter for the particular utterance is below the threshold, the utterance is flagged, and it is routed/presented over a network to a human operator. The human operator can quickly step through large numbers of these recorded utterances and listen to them (the GUI screen can be used to present and play each recording and allow the human operator to select a meaning from a list of pre-defined choices or type in free form text, depending on what the caller said). By "tuning" what responses the speech recognizer is looking for and the confidence thresholds, very high accuracy rates (higher than a computer can do without human checking) can be achieved, but with selective and semi-automated human involvement (for great cost efficiency) – an excellent trade off.

It is submitted that neither the Bartosik et al. patent nor the Papineni et al. patent either anticipates or makes obvious this invention.

The Bartosik et al. patent describes a dictation system, that is, a method/system for speaker-dependent speech recognition, that the respondent (speech recognition user) operates by speaking into a microphone attached directly to a computer. The method/system is designed to correct dictated text, by having a human operator view what was processed by the speaker-dependent speech recognition system and type in corrections. The patent also describes a method of using the corrected text to optimize the recognizer for a specific respondent's voice. It also presents ALL the text that it thinks the respondent spoke (since the system processes a continuous stream of dictated

words), not specific choices (from shorter utterances) as claimed in the present application. An application of Bartosik's system would be a doctor dictating his/her notes with respect to a patient, and then having an assistant review the automatically dictated text to correct for obvious errors. This, as opposed to present system/method – where the respondent calls in over the phone, the speech recognition application is speaker-independent (since the current application can not control who calls in), and the system/method checks specific, short responses against a list of expected responses.

For references to dictation and speaker-dependent recognition, see Bartosik et al. Fig 1, (elements 2, 42, 54) showing a microphone connected to a computer, and also column 1, lines 13 -16 (“microphone”), column 1, line 56 (“adjusting to the speaker”), column 3, lines 10, 12, 18, 41, 47, 55 (“dictation”....“microphone”....“USB”), column 6, lines 47-54 (respondent training the system on their voice), column 8, line 63 (“dictations”), column 12, line 45 (“forms a dictating machine”), column 13, lines 10-12 (“adjusted to the respective user”), column 14 lines 3-5 (microphone connected via USB to computer), among other references.

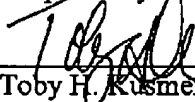
It is submitted that Papineini et al. does not overcome the deficiencies of the Bartosik et al. reference. Papineini et al. describe a “dialog manager” for controlling what prompts a speech recognition (SR) system plays next (i.e., to control the interaction). Electronic “forms” are used to select a topic or context, and “slots” in these forms to dictate what the system will ask for next. A web (or paper based) form analogy to this approach is that one might have a page on trading stocks, and on this page there would be blanks to fill in for the company name, ticker symbol, number of shares, price, etc. The forms are used to control what the system asks for (i.e., what prompts it plays) and in what order (e.g., it would ask for the company name first and fill in the company slot based on what the user said, then ask for the number of shares, etc.). The patentees are quite clear that they are always talking about automatic computer speech recognition, and there is clearly no mention of recording the calls simultaneously, using a confidence threshold to flag and route certain calls to be listened to by a human, or having a human filling out the forms for the caller/respondent. Indeed, the forms are never actually seen or used by the caller/respondent – they are just a method to deliver and control the prompts to the caller/respondent.

BEST AVAILABLE COPY

It should also be understood that the user in the presently disclosed system is a human operator who selects and plays recordings of utterances from a caller when those utterances fall outside of a predetermined limit set by a confidence parameter, and then selects response choices from a predefined list, or types in responses if they are not on a predefined list. In the Papineni disclosed system, the user is the caller – the “forms” are a tool used by the system to control what prompts are played and in what order (to “control the dialog”) and are never seen nor used by anyone beside the developer. It is clear therefore that the references could not be combined to make obvious the claimed invention of the present application.

In summary, therefore, all of the claims, claims 1, 5-13, 26, 28, 29 and 33, as now presented are believed to be patentable over the cited prior art. Applicants encourage the Examiner to call the undersigned if any questions arise, or the Examiner wishes to make suggestions to advance the prosecution of this application. Accordingly, an early and favorable action thereon, is therefore earnestly solicited. Please apply any charges or credits to deposit account 50-1133.

Respectfully submitted,



Toby H. Kusner

Registration Number 26,418
McDermott, Will & Emery
28 State Street
Boston, MA 02109-1775
Tel.: 617-535-4065
Fax: 617-535-3800

BST99 1482066-1.057622.0045